Problem 1

To enforce the necessary constraints, we introduce the Lagrangian multipliers $\lambda_1, ..., \lambda_M, \lambda_{M+1}$ and make an unconstrained maximization of

$$L = \mathbf{u}_{M+1}{}^\intercal \mathbf{S}\mathbf{u}_{M+1} + \lambda_{M+1}(1 - \mathbf{u}_{M+1}{}^\intercal \mathbf{u}_{M+1}) + \sum_{i=1}^{M} \lambda_i(\mathbf{u}_{M+1}{}^\intercal \mathbf{u}_i)$$

Now we take the derivative of L with respect to a vector $\mathbf{u}_{M+1}$ and set it to zero

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{u}_{M+1}} L = 0 \Leftrightarrow 2\mathbf{S}\mathbf{u}_{M+1} &= 2\lambda_{M+1}\mathbf{u}_{M+1} - \sum_{i=1}^{M} \lambda_i \mathbf{u}_i \\
\Leftrightarrow 2\mathbf{u}_{M+1}{}^\intercal \mathbf{S}\mathbf{u}_{M+1} &= 2\lambda_{M+1}\mathbf{u}_{M+1}{}^\intercal \mathbf{u}_{M+1} - \sum_{i=1}^{M} \lambda_i \mathbf{u}_{M+1}{}^\intercal \mathbf{u}_i && \text{(left-multiply by } \mathbf{u}_{M+1}{}^\intercal\text{)} \\
\Leftrightarrow \mathbf{u}_{M+1}{}^\intercal \mathbf{S}\mathbf{u}_{M+1} &= \lambda_{M+1} && \text{(by orthonormality of } u_i \text{ and } \mathbf{u}_{M+1}\text{)}
\end{aligned}$$

By induction step, we see that the variance in direction $\mathbf{u}_{M+1}$ is maximum when we set $\mathbf{u}_{M+1}$ equal to the eigenvector having the $(m+1)^{\text{th}}$ largest eigenvalue λ_{M+1}.

Problem 2

$$\mathbf{x_i} \sim \mathcal{N}(\mu_\mathbf{x}, \mathbf{\Sigma_x}) \qquad (\mu_\mathbf{x} = \mu,\ \mathbf{\Sigma_x} = \mathbf{W}\mathbf{W}^\intercal + \sigma^2\mathbf{I})$$

From lecture we know that $\mathbf{y_i} = \mathbf{A}\mathbf{x_i} \sim \mathcal{N}(\mu_\mathbf{y}, \mathbf{\Sigma_y})$. Let's first derive the two moments

$$\begin{aligned}
\mu_\mathbf{y} &= \mathbf{E}[y] \\
&= \mathbf{E}[\mathbf{A}\mathbf{x_i}] \\
&= \mathbf{A}\mathbf{E}[\mathbf{x_i}] \\
&= \mathbf{A}\mu_\mathbf{x}
\end{aligned}$$

$$\begin{aligned}
\mathbf{\Sigma_y} &= \mathbf{E}[(\mathbf{y} - \mu_\mathbf{y})(\mathbf{y} - \mu_\mathbf{y})^\intercal] \\
&= \mathbf{E}[(\mathbf{A}\mathbf{x} - \mathbf{A}\mu_\mathbf{x})(\mathbf{A}\mathbf{x} - \mathbf{A}\mu_\mathbf{x})^\intercal] \\
&= \mathbf{E}[(\mathbf{A}(\mathbf{x} - \mu_\mathbf{x}))(\mathbf{A}(\mathbf{x} - \mu_\mathbf{x}))^\intercal] \\
&= \mathbf{A}\mathbf{E}[(\mathbf{x} - \mu_\mathbf{x})(\mathbf{x} - \mu_\mathbf{x})^\intercal]\mathbf{A} \\
&= \mathbf{A}\mathbf{\Sigma_x}\mathbf{A}
\end{aligned}$$

By pattern matching we have for the transformed Maximum Likelihood estimates

$$\begin{aligned}
\mu_{y_{ML}} &= \mathbf{A}\mu_{ML} \\
\mathbf{\Sigma}_{y_{ML}} &= \mathbf{A}\mathbf{\Sigma}_{ML}\mathbf{A}^\intercal \\
\mathbf{W}_{y_{ML}} &= \mathbf{A}\mathbf{W}_{ML}
\end{aligned}$$

Problem 3

Let $\mathbf{x} \in \mathbb{R}^5$ hold the movie ratings given by Leslie.
By the SVD projection in concept space we have

$$\mathbf{v} \cdot \mathbf{x} = [1.74, 2.84]^\intercal$$

By SVD decomposition and reconstruction of the input using the projected space we have

$$[1.74, 2.84]^\intercal \cdot \mathbf{v}^\intercal = [1.0092, 1.0092, 1.0092, 2.0164, 2.0164]^\intercal$$

E.g. we can predict Leslie to rate Titanic movie with 2.0164.

Problem 4

See below.